

FORM PTO-1390 (Modified)  
(REV 11-98)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER

## TRANSMITTAL LETTER TO THE UNITED STATES

PL-9830

DESIGNATED/ELECTED OFFICE (DO/EO/US)

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR

CONCERNING A FILING UNDER 35 U.S.C. 371

To be assigned 09/787228

INTERNATIONAL APPLICATION NO.

PCT/GB99/03081

INTERNATIONAL FILING DATE

14 September 1999

PRIORITY DATE CLAIMED

14 September 1998

TITLE OF INVENTION

Gene and Protein Libraries and Methods Relating Thereto

APPLICANT(S) FOR DO/EO/US

Anna Hine, Leonie Morgan, Albert Santos, and David Palfrey

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371 (c) (2))
  - a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☒ A copy of the International Search Report (PCT/ISA/210).
8. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
  - a. ☒ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
9. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
10. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371 (c)(4)).
11. ☒ A copy of the International Preliminary Examination Report (PCT/IPEA/409).
12. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).

## Items 13 to 20 below concern document(s) or information included:

13. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
14. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
15. ☒ A **FIRST** preliminary amendment.
16. ☐ A **SECOND** or **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.
19. ☒ Certificate of Mailing by Express Mail
20. ☒ Other items or information:

Copy of the International Application as Published by the International Bureau

Copy of this transmittal letter for charging purposes

A Submission of Nucleotide and/or Amino Acid Sequence Disclosures with CFR and printed copies

Return Postcard

U.S. APPLICATION NO. (IF KNOWN, SEE 37 CFR 1.53) <b>To be assigned</b> <b>09/787228</b>	INTERNATIONAL APPLICATION NO. <b>PCT/GB99/03081</b>	ATTORNEY'S DOCKET NUMBER <b>PL-9830</b>
--	--	--

21. The following fees are submitted:				<b>CALCULATIONS PTO USE ONLY</b>	
<b>BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)) :</b>					
<input type="checkbox"/> Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO .....				\$1,000.00	
<input checked="" type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO .....				\$860.00	
<input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO .....				\$710.00	
<input type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) .....				\$690.00	
<input type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) .....				\$100.00	
<b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>				<b>\$860.00</b>	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492 (e)).				<b>\$0.00</b>	
<b>CLAIMS</b>	<b>NUMBER FILED</b>	<b>NUMBER EXTRA</b>	<b>RATE</b>		
Total claims	27 - 20 =	7	x \$18.00	<b>\$126.00</b>	
Independent claims	4 - 3 =	1	x \$80.00	<b>\$80.00</b>	
Multiple Dependent Claims (check if applicable).			<input type="checkbox"/>	<b>\$0.00</b>	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				<b>\$1,066.00</b>	
Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28) (check if applicable).				<input type="checkbox"/> <b>\$0.00</b>	
<b>SUBTOTAL =</b>				<b>\$1,066.00</b>	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492 (f)).				<b>\$0.00</b>	
<b>TOTAL NATIONAL FEE =</b>				<b>\$1,066.00</b>	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31) (check if applicable).				<input type="checkbox"/> <b>\$0.00</b>	
<b>TOTAL FEES ENCLOSED =</b>				<b>\$1,066.00</b>	
				Amount to be refunded	\$
				charged	\$

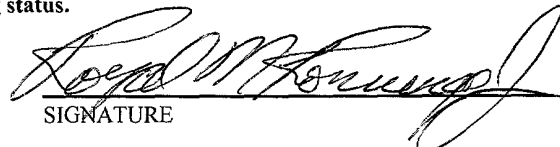
- ☐ A check in the amount of \_\_\_\_\_ to cover the above fees is enclosed.
- ☒ Please charge my Deposit Account No. **500-588** in the amount of **\$1,066.00** to cover the above fees.  
A duplicate copy of this sheet is enclosed.
- ☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. **500-588** A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

Royal N. Ronning, Jr.  
Amersham Pharmacia Biotech, Inc.  
800 Centennial Avenue  
Piscataway, New Jersey 08855

(732) 457-8423

  
SIGNATURE

Royal N. Ronning, Jr.

NAME

32,529

REGISTRATION NUMBER

March 14, 2001

DATE

09/787228

JC02 Rec'd PCT/PTO 1 4 MAR 2001

**CERTIFICATE OF MAILING**

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents Washington, D.C. 20231.

Melissa Leck  
Type or print name

  
Signature

March 14, 2001  
Date

Application No.:	To be assigned
Filing Date:	March 14, 2001
Inventors:	A. Hine, et al.
Title:	Gene and Protein Libraries and Methods Relating Thereto
File Reference No.:	PL-9830

**Assistant Commissioner for Patents  
United States Patent and Trademark Office  
Washington, D.C. 20231**

SUBMISSION OF NUCLEOTIDE AND/OR AMINO ACID SEQUENCE DISCLOSURES

Dear Sirs:

In connection with the prosecution of the captioned application, which claims priority to international application number PCT/GB99/03081 filed September 14, 1999, Applicants submit the following items:

- 1) An initial computer readable form (CFR) copy of the sequence listing and an initial paper copy of the sequence listing, both generated on an IBM computer using PatentIn, Version 2.1 Software.
- 2) Required statements:
  - A. Applicants aver that the sequence listings do not include any new matter which goes beyond the content of the application as filed.
  - B. Applicants aver that that sequence listings contained on the diskette are identical to those contained on the paper copy.

09/787228 071401

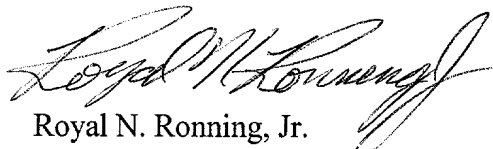
09/787228

JC02 Rec'd PCT/PTO 14 MAR 2001

Applicants simultaneously submit a First Preliminary Amendment instructing that the attached sequence listings be inserted as a part of the specification. Please direct any remaining issues to Applicant's counsel at the telephone number provided below.

Amersham Pharmacia Biotech, Inc.  
P.O. Box 1327  
800 Centennial Avenue  
Piscataway, New Jersey 08855-1327  
(732)457-8423 (voice)  
(732)457-8463 (facsimile)

Respectfully submitted,

  
Royal N. Ronning, Jr.  
Reg. No. 32,529

106120-822/3/60

PL-9830

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Application of: A. Hine, et al. Group Art Unit: To be assigned  
Serial Number: To be assigned Examiner: To be assigned  
Filing Date: 14 March 2001  
Title: Gene and Protein Libraries and Methods Relating Thereto

**FIRST PRELIMINARY AMENDMENT**

Honorable Assistant Commissioner of Patents  
Box New Patent Application  
Washington, D.C. 20231

Sir:

Please consider the following amendments and remarks in connection with the prosecution of the captioned application, which claims priority to international application number PCT/GB99/03081 filed September 14, 1999.

**IN THE SPECIFICATION**

At the end of the written description, before the claims, please insert the Sequence Listing attached hereto.

09/787228-0715001

**IN THE CLAIMS**

Please amend the claims as follows:

**[CLAIMS]**

**WHAT IS CLAIMED IS:**

1. (amended) A set of libraries of genes which code for proteins which are capable of specific binding interactions with a specific binding partner by amino acid residues at at least two specified positions including a first specified position and at least one other specified position, which set of libraries [consists of]comprises:

a) 6 to 20 libraries in which each library has a triplet that codes for at least one but less than 20 amino acids at said first specified position, and is [randomised]randomized at the or each triplet coding for the said at least one other specified position, the arrangement being such that interactions of the proteins coded for by the said 6 to 20 libraries with a specific binding partner identifies a triplet that codes for an amino acid at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is [randomised]randomized and a different one of said at least one other specified positions has a triplet that codes for at least one but less than 20 amino acids.

2. (amended) The set of libraries of genes as claimed in claim 1, which set of libraries [consists of]comprises:

a) 12 libraries in which each library has a triplet that codes for one or several but less than 20 amino acids at the said first determined position, the triplets being as shown in Table 1 or Table 2, and

b) 12 libraries of corresponding design for each of the said one or more other determined positions.

3. (amended) The set of libraries of genes as claimed in claim 1[ or claim 2], wherein the genes code for zinc fingers.

4. (amended) The set of libraries of genes as claimed in claim 3, which set [consists of]comprises 36 libraries in three groups of 12 libraries which code for amino acids at the -1 and +3 and +6 positions respectively.

5. (amended) The set of libraries of genes as claimed in claim 3[ or claim 4], wherein each gene codes for a protein comprising [3]three zinc fingers.

6. (amended) The set of libraries of genes as claimed in claim 5, wherein each gene codes for a protein having the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

where X is any amino acid.

7. (amended) A set of libraries of proteins, which proteins are capable of specific binding interactions with a specified binding partner by amino acid residues at at least one specified position including a first specified position and at least one other specified position, which set of libraries [consists of]comprises:

a) 6 to 20 libraries in which each library has at least one but less than 20 amino acid residues at the said first specified position and is [randomised]randomized at the said at least one other determined position, the arrangement being such that interaction of the 6 to 20 libraries with a specific binding partner identifies an amino acid residue at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is [randomised]randomized and a different amino acid is present at at least one other specified position.

8. (amended) The set of libraries of proteins as claimed in claim 7, which set of libraries [consists of]comprises:

a) 20 libraries in which each library has one specified amino acid residue at the said first determined position and is [randomised]randomized at the said one or more other determined positions, and

b) 20 libraries of corresponding design for each of the said one or more other determined positions.

9. (amended) The set of libraries or proteins as claimed in claim 7[ or claim 8], wherein the proteins are zinc fingers.

10. (amended) The set of libraries of proteins as claimed in claim 7, which set [consists of]comprises 60 libraries in three groups of 20 libraries with specified amino acids at the -1 and +3 and +6 positions respectively.

11. (amended) The set of libraries of proteins as claimed in claim 9[ or claim 10], wherein each protein comprises three zinc fingers.

12. (amended) The set of libraries of proteins as claimed in claim 11, wherein each protein has the sequence (SEQ ID NO: 2)

TGEKPYKCPECGKSFSXKSXLVXHQRTH

TGEKPYKCPECGKSFSXKSXLVXHQRTH

TGEKPYKCPECGKSFSXKSXLVXHQRTH

where X is any amino acid.

13. (amended) A set of libraries of genes which code for the set of libraries of proteins defined in [any one of claims 7 to 12]claim 7.

14. (amended) A method of identifying a protein which interacts with a specific binding partner, which method comprises providing a set of libraries of proteins as defined in [any one of claims



7 to 12]claim 7, incubating the specific binding partner with each library of the set, observing specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner.

19. (amended) The method as claimed in claim 18, wherein the sets of libraries of proteins are [immobilised]immobilized on scintillation proximity assay surfaces and the specific binding partner is radiolabelled.

20. (amended) The method of claim 18[ or claim 19], wherein after incubation the scintillation proximity assay surfaces are washed to distinguish stronger specific binding interactions from weaker ones.

23. (amended) The method as claimed in claim 21,[ or claim 22] wherein after incubation the binding interactions are washed to distinguish stronger specific binding interactions from weaker ones.

24. (amended) A protein having the sequence (SEQ ID NO: 1)

T G E K P Y K C P E C G K S F S [.]K K S [+]H L V [\$]A H Q R T H

T G E K P Y K C P E C G K S F S [.]K K S [+]H L V [\$]A H Q R T H

T G E K P Y K C P E C G K S F S [.]K K S [+]H L V [\$]A H Q R T H.

26. (amended) A method of constructing [randomised]randomized gene libraries in which the number of genes is the same as the number of encoded proteins and which contain no termination codons at the predetermined positions of [randomisation]randomization, the method comprising the steps of:

a) providing a template oligonucleotide which is fully [randomised]randomized at predetermined codon positions;

b) for each predetermined codon position providing a pool of selection oligonucleotides, wherein each member of said pool contains a different codon selected from the group [consisting of]comprising:

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA,  
GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

at the predetermined codon position;

- c) selecting one or more selection oligonucleotides from each pool in order to encode the required gene or library;
- d) allowing the ligated selected oligonucleotides from each pool to [hybridise]hybridize with the template oligonucleotide;
- e) forming one or more constructs by ligating the [hybridised]hybridized selection oligonucleotides together;
- f) removing a region from a gene of interest corresponding to the [hybridised]hybridized product from step e);
- g) forming a gene library or genes by ligating the products from step e) into the said gene of interest wherein the said gene of interest is contained within a suitable expression vector.

27. (amended) A method of producing proteins encoded by the [randomised]randomized gene libraries of claim 26 comprising the steps of:

- a) transforming a suitable host cell with the gene or gene library of claim [27]26 construct;
- b) expressing the genes to form proteins;
- c) purifying the proteins.

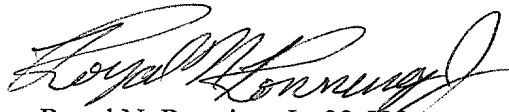
**REMARKS**

Claims 1-27 are pending in the instant application. Applicants have amended claims 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 19, 20, 23, 24, 26, and 27 to more fully conform with U.S. practice and to delete multiple dependencies. A copy of the marked up claims showing the amendments, as well as a clean copy of the claims encompassing the amendments, is attached hereto.

Applicants respectfully assert that all amendments are fairly based on the specification, and respectfully request their entry.

Applicants believe that the claims, as amended, are in allowable form, and earnestly solicit the allowance of claims 1-27.

Respectfully submitted,



Royal N. Ronning, Jr. 32,529  
Attorney for Applicants

Amersham Pharmacia Biotech, Inc.  
800 Centennial Avenue  
P. O. Box 1327  
Piscataway, NJ 08855-1327

Tel: (732) 457-8423  
Fax: (732) 457-8463

H:\DEPT\Law\_Dept\IP Dept\US Patent Forms PDF & Word\Amedments\First Preliminary Amendment-PL9830 (3-14-01).doc

**Amended Claims (marked up copy showing amendment(s))**

**[CLAIMS]**

**WHAT IS CLAIMED IS:**

1. (amended) A set of libraries of genes which code for proteins which are capable of specific binding interactions with a specific binding partner by amino acid residues at at least two specified positions including a first specified position and at least one other specified position, which set of libraries [consists of]comprises:

a) 6 to 20 libraries in which each library has a triplet that codes for at least one but less than 20 amino acids at said first specified position, and is [randomised]randomized at the or each triplet coding for the said at least one other specified position, the arrangement being such that interactions of the proteins coded for by the said 6 to 20 libraries with a specific binding partner identifies a triplet that codes for an amino acid at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is [randomised]randomized and a different one of said at least one other specified positions has a triplet that codes for at least one but less than 20 amino acids.

2. (amended) The set of libraries of genes as claimed in claim 1, which set of libraries [consists of]comprises:

a) 12 libraries in which each library has a triplet that codes for one or several but less than 20 amino acids at the said first determined position, the triplets being as shown in Table 1 or Table 2, and

b) 12 libraries of corresponding design for each of the said one or more other determined positions.

3. (amended) The set of libraries of genes as claimed in claim 1 [ or claim 2], wherein the genes code for zinc fingers.

4. (amended) The set of libraries of genes as claimed in claim 3, which set [consists of]comprises 36 libraries in three groups of 12 libraries which code for amino acids at the -1 and +3 and +6 positions respectively.

5. (amended) The set of libraries of genes as claimed in claim 3[ or claim 4], wherein each gene codes for a protein comprising [3]three zinc fingers.

6. (amended) The set of libraries of genes as claimed in claim 5, wherein each gene codes for a protein having the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

where X is any amino acid.

7. (amended) A set of libraries of proteins, which proteins are capable of specific binding interactions with a specified binding partner by amino acid residues at at least one specified position including a first specified position and at least one other specified position, which set of libraries [consists of]comprises:

a) 6 to 20 libraries in which each library has at least one but less than 20 amino acid residues at the said first specified position and is [randomised]randomized at the said at least one other determined position, the arrangement being such that interaction of the 6 to 20 libraries with a specific binding partner identifies an amino acid residue at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is [randomised]randomized and a different amino acid is present at at least one other specified position.

8. (amended) The set of libraries of proteins as claimed in claim 7, which set of libraries [consists of]comprises:

a) 20 libraries in which each library has one specified amino acid residue at the said first determined position and is [randomised]randomized at the said one or more other determined positions, and

b) 20 libraries of corresponding design for each of the said one or more other determined positions.

9. (amended) The set of libraries or proteins as claimed in claim 7[ or claim 8], wherein the proteins are zinc fingers.

10. (amended) The set of libraries of proteins as claimed in claim 7, which set [consists of]comprises 60 libraries in three groups of 20 libraries with specified amino acids at the -1 and +3 and +6 positions respectively.

11. (amended) The set of libraries of proteins as claimed in claim 9[ or claim 10], wherein each protein comprises three zinc fingers.

12. (amended) The set of libraries of proteins as claimed in claim 11, wherein each protein has the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

where X is any amino acid,

13. (amended) A set of libraries of genes which code for the set of libraries of proteins defined in [any one of claims 7 to 12]claim 7.

14. (amended) A method of identifying a protein which interacts with a specific binding partner, which method comprises providing a set of libraries of proteins as defined in [any one of claims 7 to 12]claim 7, incubating the specific binding partner with each library of the set, observing

specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner.

19. (amended) The method as claimed in claim 18, wherein the sets of libraries of proteins are [immobilised]immobilized on scintillation proximity assay surfaces and the specific binding partner is radiolabelled.

20. (amended) The method of claim 18[ or claim 19], wherein after incubation the scintillation proximity assay surfaces are washed to distinguish stronger specific binding interactions from weaker ones.

23. (amended) The method as claimed in claim 21,[ or claim 22] wherein after incubation the binding interactions are washed to distinguish stronger specific binding interactions from weaker ones.

24. (amended) A protein having the sequence (SEQ ID NO: 1)

T G E K P Y K C P E C G K S F S [.]K K S [ + ] H L V [ \$ ] A H Q R T H  
T G E K P Y K C P E C G K S F S [.]K K S [ + ] H L V [ \$ ] A H Q R T H  
T G E K P Y K C P E C G K S F S [.]K K S [ + ] H L V [ \$ ] A H Q R T H.

26. (amended) A method of constructing [randomised]randomized gene libraries in which the number of genes is the same as the number of encoded proteins and which contain no termination codons at the predetermined positions of [randomisation]randomization, the method comprising the steps of:

- a) providing a template oligonucleotide which is fully [randomised]randomized at predetermined codon positions;
- b) for each predetermined codon position providing a pool of selection oligonucleotides, wherein each member of said pool contains a different codon selected from the group [consisting of]comprising:

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA,

GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

at the predetermined codon position;

- c) selecting one or more selection oligonucleotides from each pool in order to encode the required gene or library;
- d) allowing the ligated selected oligonucleotides from each pool to [hybridise]hybridize with the template oligonucleotide;
- e) forming one or more constructs by ligating the [hybridised]hybridized selection oligonucleotides together;
- f) removing a region from a gene of interest corresponding to the [hybridised]hybridized product from step e);
- g) forming a gene library or genes by ligating the products from step e) into the said gene of interest wherein the said gene of interest is contained within a suitable expression vector.

27. (amended) A method of producing proteins encoded by the [randomised]randomized gene libraries of claim 26 comprising the steps of:

- a) transforming a suitable host cell with the gene or gene library of claim [27]26 construct;
- b) expressing the genes to form proteins;
- c) purifying the proteins.



**Clean copy of claims showing amendment(s)**

**WHAT IS CLAIMED IS:**

1. (amended) A set of libraries of genes which code for proteins which are capable of specific binding interactions with a specific binding partner by amino acid residues at at least two specified positions including a first specified position and at least one other specified position, which set of libraries comprises:

a) 6 to 20 libraries in which each library has a triplet that codes for at least one but less than 20 amino acids at said first specified position, and is randomized at the or each triplet coding for the said at least one other specified position, the arrangement being such that interactions of the proteins coded for by the said 6 to 20 libraries with a specific binding partner identifies a triplet that codes for an amino acid at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is randomized and a different one of said at least one other specified positions has a triplet that codes for at least one but less than 20 amino acids.

2. (amended) The set of libraries of genes as claimed in claim 1, which set of libraries comprises:

a) 12 libraries in which each library has a triplet that codes for one or several but less than 20 amino acids at the said first determined position, the triplets being as shown in Table 1 or Table 2, and

b) 12 libraries of corresponding design for each of the said one or more other determined positions.

3. (amended) The set of libraries of genes as claimed in claim 1, wherein the genes code for zinc fingers.

4. (amended) The set of libraries of genes as claimed in claim 3, which set comprises 36 libraries in three groups of 12 libraries which code for amino acids at the -1 and +3 and +6 positions respectively.

5. (amended) The set of libraries of genes as claimed in claim 3, wherein each gene codes for a protein comprising three zinc fingers.

6. (amended) The set of libraries of genes as claimed in claim 5, wherein each gene codes for a protein having the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

where X is any amino acid.

7. (amended) A set of libraries of proteins, which proteins are capable of specific binding interactions with a specified binding partner by amino acid residues at at least one specified position including a first specified position and at least one other specified position, which set of libraries comprises:

a) 6 to 20 libraries in which each library has at least one but less than 20 amino acid residues at the said first specified position and is randomized at the said at least one other determined position, the arrangement being such that interaction of the 6 to 20 libraries with a specific binding partner identifies an amino acid residue at the said first specified position that takes part in the specific binding interaction, and

b) 6 to 20 libraries in each of which libraries said first specified position is randomized and a different amino acid is present at at least one other specified position.

8. (amended) The set of libraries of proteins as claimed in claim 7, which set of libraries comprises:

- a) 20 libraries in which each library has one specified amino acid residue at the said first determined position and is randomized at the said one or more other determined positions, and
- b) 20 libraries of corresponding design for each of the said one or more other determined positions.

9. (amended) The set of libraries or proteins as claimed in claim 7, wherein the proteins are zinc fingers.

10. (amended) The set of libraries of proteins as claimed in claim 7, which set comprises 60 libraries in three groups of 20 libraries with specified amino acids at the -1 and +3 and +6 positions respectively.

11. (amended) The set of libraries of proteins as claimed in claim 9, wherein each protein comprises three zinc fingers.

12. (amended) The set of libraries of proteins as claimed in claim 11, wherein each protein has the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H  
T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H  
T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

where X is any amino acid.

13. (amended) A set of libraries of genes which code for the set of libraries of proteins defined in claim 7.

14. (amended) A method of identifying a protein which interacts with a specific binding partner, which method comprises providing a set of libraries of proteins as defined in claim 7, incubating

the specific binding partner with each library of the set, observing specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner.

15. The method as claimed in claim 14, wherein the specific binding partner is a polynucleotide.

16. The method as claimed in claim 14, wherein the specific binding interactions are observed by radiometric or luminescent assay.

17. The method as claimed in claim 14, wherein the specific binding interactions are observed by imaging means.

18. The method as claimed in claim 14, wherein the specific binding interactions are observed by scintillation proximity assay.

19. (amended) The method as claimed in claim 18, wherein the sets of libraries of proteins are immobilized on scintillation proximity assay surfaces and the specific binding partner is radiolabelled.

20. (amended) The method of claim 18, wherein after incubation the scintillation proximity assay surfaces are washed to distinguish stronger specific binding interactions from weaker ones.

21. The method as claimed in claim 14, wherein the specific binding interactions are observed by colorimetric means.

22. The method as claimed in claim 21, wherein the specific binding partner is biotinylated and the specific binding interaction is detected using a signal generating streptavidin conjugate.

23. (amended) The method as claimed in claim 21, wherein after incubation the binding interactions are washed to distinguish stronger specific binding interactions from weaker ones.

24. (amended) A protein having the sequence (SEQ ID NO: 1)

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H.

25. A gene which codes for the protein of claim 24.

26. (amended) A method of constructing randomized gene libraries in which the number of genes is the same as the number of encoded proteins and which contain no termination codons at the predetermined positions of randomization, the method comprising the steps of:

a) providing a template oligonucleotide which is fully randomized at predetermined codon positions;

b) for each predetermined codon position providing a pool of selection oligonucleotides, wherein each member of said pool contains a different codon selected from the group comprising:

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA,  
GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

at the predetermined codon position;

c) selecting one or more selection oligonucleotides from each pool in order to encode the required gene or library;

d) allowing the ligated selected oligonucleotides from each pool to hybridize with the template oligonucleotide;

e) forming one or more constructs by ligating the hybridized selection oligonucleotides together;

f) removing a region from a gene of interest corresponding to the hybridized product from step e);

g) forming a gene library or genes by ligating the products from step e) into the said gene of interest wherein the said gene of interest is contained within a suitable expression vector.

27. (amended) A method of producing proteins encoded by the randomized gene libraries of claim 26 comprising the steps of:

- a) transforming a suitable host cell with the gene or gene library of claim 26 construct;
- b) expressing the genes to form proteins;
- c) purifying the proteins.

## GENE AND PROTEIN LIBRARIES AND METHODS RELATING THERETO

5 Introduction

Naturally occurring proteins are capable of specific binding interactions with other proteins and other molecules. It is well known that such proteins can be used as scaffolds and specific amino acid residues changed in order to improve binding properties. The changes required can  
10 be determined by combinatorial chemistry means. The subject is reviewed by Per-Ake Nygren and Mathias Uhlen in Curr. Opin. Struct. Biol. (1997) 7, 463-469, who list cyclic peptides, immunoglobulin-like scaffolds, bacterial receptors, DNA-binding proteins and protease inhibitors as examples of protein scaffolds. The authors conclude that, starting from a suitable  
15 protein domain, the use of a combinatorial approach coupled with powerful selection or screening strategies can be used to obtain novel proteins capable of binding a desired target molecule. But the selection or screening strategies can be difficult. It is this problem that is addressed by the present invention.

20 Zinc fingers are examples of protein scaffolds of the kind described. Zinc fingers are protein motifs ("mini-domains") which interact with double-stranded DNA (some also bind RNA). This interaction is dependent on DNA sequence, thus the interaction is termed to be sequence-specific. The interaction between the zinc finger and its target  
25 DNA sequence is modular: one zinc finger recognises three bases of DNA. Basic rules concerning the interaction were determined early on by structural studies (both X-ray crystallography and NMR spectroscopy) of zinc finger-DNA complexes. In essence, three residues (amino acids) within the zinc finger make base-specific contacts with the DNA. These  
30 three residues differ greatly between different zinc fingers, allowing a limited repertoire of different DNA sequences to be recognised. Early

09787228.071901

mutagenesis experiments determined that if these variable residues are changed, a different DNA sequence may be recognised. (A fourth residue sometimes contributes to DNA recognition, but this residue is well-conserved between different zinc finger proteins). In practice then, the zinc finger may be viewed as a molecular scaffold, which orientates the three variable residues suitably to enable them to make base-specific contacts with the DNA.

It would be most advantageous to have available a zinc finger to bind each trinucleotide (3 bases) of dsDNA. Initial attempts to achieve this goal centred on the structure-based design of novel zinc finger proteins. Since 1994 however, several groups have employed combinatorial libraries of zinc finger proteins and/or target DNA sequences to identify novel zinc fingers which bind to the required DNA sequences

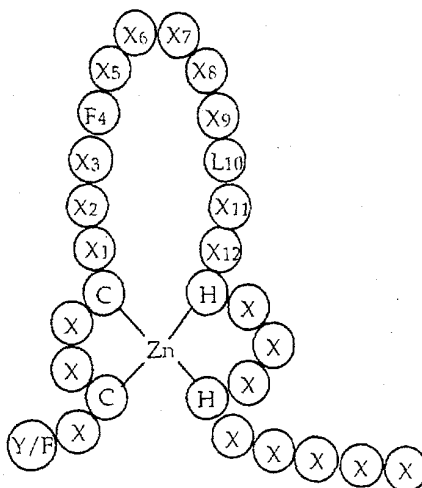
One such technique has been developed by Choo and Klug and is described in WO 96/06166 and in PNAS, 91, 11163-11167 and 11168-11172 (1994). A single library of zinc finger genes was constructed. The library was based on a naturally occurring zinc finger protein, Zif 268, which contains three zinc fingers. Only the central finger was randomised at seven positions. The library of genes was cloned as a fusion to the fd phage gene pIII. When expressed, a library of bacteriophage resulted, in which each bacteriophage displayed a randomised zinc finger protein on its surface. In a first stage assay, this library was incubated with a target DNA molecule, and individual clones that bound to the target were purified and sequenced. In a second stage assay, each of those clones selected was incubated with a variety of related DNA sequences in order to further investigate its binding properties. The technique is subject to some inherent disadvantages:

- Deconvolution is not addressed – purification is inherent in the method. The assay results in a pool of a bacteriophage. For identification purposes, each member of that pool must be cultured independently and its DNA sequenced.



- The present invention addresses these shortcomings.
- Zinc fingers are small protein motifs. They form parts of  
15 larger proteins, but perform their specific function within those proteins.  
Zinc fingers exist in tandem arrays: proteins containing between 2 and 37  
different zinc fingers have been identified.

In two dimensions, a single zinc finger appears as follows:



In this diagram, each circle represents a single amino acid

residue.

The zinc finger is so stable that its structure is unaffected by the replacement of virtually all residues marked "X" with alanine (Michael *et al*, PNAS 89, 4796-4800, 1992). Spaced correctly (as above) the

5 following requirements are all that are necessary for the formation of a zinc finger:

- The 2 cysteine (C) residues
  - The 2 histidine (H) residues
  - The zinc ion (Zn), which is co-ordinated (bound) by the C and
- 10 H residues
- Three hydrophobic residues: tyrosine/phenylalanine (Y/F); phenylalanine (F4); leucine (L10).

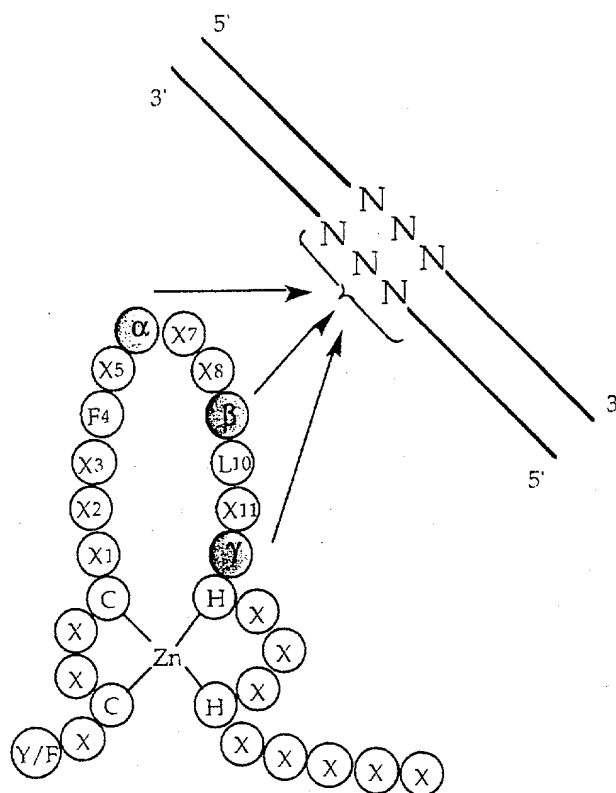
Zinc fingers bind to nucleic acids - either DNA or RNA. In nature, zinc fingers usually form part of transcription factors, but in the

15 laboratory, it is possible to work with them independently from the rest of these proteins. The zinc finger exemplified herein binds to double-stranded DNA. One zinc finger binds to three bases of DNA (a trinucleotide).

Several zinc fingers are usually linked in tandem. Most frequently, three zinc fingers interact with successive trinucleotides, which

20 means that altogether, the three zinc fingers will interact with (recognise) a specific 9 base pair (bp) sequence of DNA. Each zinc finger will recognise a specific trinucleotide. However, nature has only provided a limited repertoire of zinc fingers, so the number of 9 base pair sequences which can be recognised is very limited.

25 The mechanism of DNA recognition is sequence-specific and surprisingly simple. Three residues (amino acids) within the zinc finger make contacts (hydrogen bonds or Van de Waal's interactions, for example) with three bases of DNA. Most of these contacts are with one strand of the DNA.



5.

10

Number of trinucleotides NNN =  $4 \times 4 \times 4 = 64$

(A,C,G or T)

15

code. A problem (addressed by this invention) is to develop such a code.

This invention involves applying the principles of combinatorial chemistry to the problem. The key to any combinatorial system (whether biological, chemical or any other system) is deconvolution:

5 the identification of an active substituent from within a mixture. The key to discovering an optimal zinc finger for each trinucleotide is to identify the optimum combinations of residues  $\alpha$ ,  $\beta$  and  $\gamma$ . There will be an optimum combination of  $\alpha$ ,  $\beta$  and  $\gamma$  for each trinucleotide. By using multiple libraries of zinc fingers, with highly controlled overlap between the libraries,

10 deconvolution can be achieved without purification.

### The Invention

In one aspect the invention provides a set of libraries of genes which code for proteins which are capable of specific binding

15 interactions by virtue of amino acid residues at two or more determined positions including a first determined position and one or more other determined positions, which set of libraries consists of:

a) 6 to 20 libraries in which each library has a triplet that codes for one or several but less than 20 amino acids at the said first determined

20 position, and is randomised at the triplet or triplets coding for the said one or more other determined positions, the arrangement being such that interactions of the proteins coded for by the said 6 to 20 libraries with a specific binding partner identifies a triplet that codes for an amino acid at the said first determined position that takes part in the specific binding

25 interaction, and

b) 6 to 20 libraries of corresponding design for each of the said one or more other determined positions.

In another aspect the invention provides a method of constructing randomised gene libraries in which the number of genes is the

30 same as the number of encoded proteins and which contain no termination codons at the predetermined positions of randomisation, the method

- 7 -

comprising the steps of:

a) providing a template oligonucleotide which is fully randomised at one or more predetermined codon positions;

b) for each predetermined codon position providing a pool of  
5 selection oligonucleotides, wherein each member of said pool contains a different codon selected from the group consisting of

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA,  
GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

10

at the predetermined codon position;

c) selecting one or more selection oligonucleotides from each pool in order to encode the required gene or library;

d) allowing the selected selection oligonucleotides from each  
15 pool to hybridise with the template oligonucleotide;

e) forming one or more constructs by ligating the hybridised selection oligonucleotides together;

f) removing a region from a gene of interest corresponding to the hybridised product from step e);

20

g) forming a gene or library of genes by ligating the products from step e) into the said gene of interest wherein the said gene of interest is contained within a suitable expression vector. A preferred method of selecting one or more selection oligonucleotides from each pool in order to encode the required gene or library at step c), is to select the selection  
25 oligonucleotides according to randomisation strategy B, described herein.

A method of producing proteins encoded by these randomised gene libraries is also provided by the invention and comprises the steps of:

a) transforming a suitable host cell with a gene or gene library construct;

30

b) expressing the genes to form proteins;

c) purifying the proteins.

FOOTNOTES: 071904

Suitable host cells, gene expression methods and purification protocols for carrying out this method are known in the art.

In another aspect the invention provides a set of libraries of proteins, which proteins are capable of specific binding interactions by virtue of amino acid residues at two or more determined positions including a first determined position and one or more other determined positions, which set of libraries consists of:

- a) 6 to 20 libraries in which each library has one or several but less than 20 amino acid residues at the said first determined position and is randomised at the said one or more other determined positions, the arrangement being such that interaction of the 6 to 20 libraries with a specific binding partner identifies an amino acid residue at the said first determined position that takes part in the specific binding interaction, and
- b) 6 to 20 libraries of corresponding design for each of the said one or more other determined positions.

In another aspect the invention provides a method of identifying a protein which interacts with a specific binding partner, which method comprises providing a set of libraries of proteins as defined, incubating the specific binding partner with each library of the set, observing specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner. Preferably, as discussed in more detail below, this method may be performed using radiometric or non-radiometric detection means, for example scintillation detection, luminescence, for example fluorescence, detection, colorimetric detection, or imaging, by methods known in the art.

A library of compounds (e.g. genes or proteins) consists of a plurality of compounds which are all different but which have some characteristic in common. The compounds of the library may be presented either separate or together, in solution or solid phase. In a set of libraries, the compounds of any one library have some characteristic in common but which differentiates them from the compound of each other library of the

set.

A specific binding interaction of a protein with another molecule (the specific binding partner) is an interaction mediated by a specified amino acid residue at one or more usually several positions in the protein molecule. The specific binding partner is usually though not necessarily a polymeric molecule, e.g. a nucleic acid (DNA or RNA) or another protein.

In relation to proteins, the statement that a library is randomised at a determined position is herein used to mean that the library contains a random mixture of all or almost all possible amino acid residues. We say "almost all" because there might be a special reason for omitting one residue e.g. Cys, or a few amino acid residues. In relation to genes, the statement that a triplet is randomised is herein used to indicate a triplet NNN (where N is any nucleotide) or a triplet that is capable of coding for all or almost all the amino acids.

The term protein is herein used to encompass any chain of two or more amino acid residues.

The term polynucleotide is herein used to encompass any chain of three or more nucleotide residues, single-stranded or double-stranded DNA or RNA.

The experimental section below describes a set of libraries of zinc finger genes which code for a set of libraries of zinc finger proteins, which are used to identify specific zinc fingers which interact with specific polynucleotides. But the invention is more broadly applicable. It is in principle possible to make a set of libraries of any protein which undergoes a specific binding interaction, using that protein as a scaffold to vary specific amino acid residues. It is in principle possible to make a set of libraries of genes coding for such a set of protein libraries. And it is possible to use such a set of protein libraries to investigate any specific binding interaction, e.g. where the specific binding partner is a polynucleotide or another protein or a different molecule. It may be noted

00707228-071901

that zinc fingers may be capable of undergoing specific binding interactions, not only with polynucleotides, but also with other proteins.

It is convenient to control the overlap between libraries of a set of protein libraries by controlling the DNA sequences of the genes which code for the proteins. Thus, to make a library of zinc finger proteins, a library of zinc finger genes is first made. For convenience in relation to what follows we quote the genetic code which relates the identities of codons to the amino acids which they specify.

		2nd base					
		A C G T					
1st base	A	Lys	Thr	Arg	Ile	A C G T	3rd base
		Asn	Thr	Ser	Ile		
		Lys	Thr	Arg	Met		
		Asn	Thr	Ser	Ile		
	C	Gln	Pro	Arg	Leu	A C G T	
		His	Pro	Arg	Leu		
		Gln	Pro	Arg	Leu		
		His	Pro	Arg	Leu		
	G	Glu	Ala	Gly	Val	A C G T	
		Asp	Ala	Gly	Val		
		Glu	Ala	Gly	Val		
		Asp	Ala	Gly	Val		
	T	STOP	Ser	STOP	Leu	A C G T	
		Tyr	Ser	Cys	Phe		
		STOP	Ser	Trp	Leu		
		Tyr	Ser	Cys	Phe		

10

Thus for example a codon with multiple degeneracy, e.g. ANN comprises 16 different triplets and codes for seven different amino acids namely Lys, Asn, Thr, Arg, Ser, Ile and Met.

While it is possible in principle to use as few as six libraries of genes to identify a particular amino acid residue, it is in practice convenient to use twelve such libraries in groups of four, wherein libraries 1 to 4 identify the first nucleotide of a triplet, libraries 5 to 8 identify the second nucleotide of the triplet, and libraries 9 to 12 identify the third nucleotide of the triplet which codes for the amino acid. In this arrangement it is preferable that only one of libraries 1 to 4 (and correspondingly only one of libraries 5 to 8 and only one of libraries 9 to 12) codes for any particular

20



amino acid. These considerations give rise to various possible sets of 12 libraries of which one is shown in the following Table 1.

**Table 1**

5

Library	Residue	Codon	Amino Acids Specified
1	$\alpha$	A <sup>A</sup> C <sub>T</sub> N	Lys Asn Thr Ile Met
2	$\alpha$	C <sup>A</sup> C <sub>G</sub> N	Gln His Pro Arg
3	$\alpha$	G N N	Au Asp Ala Gly Val
4	$\alpha$	T N N	Tyr Ser Cys Trp Leu Phe
5	$\alpha$	N A N	Lys Asn Gln His Glu Asp Tyr
6	$\alpha$	N C N	Thr Pro Ala Ser
7	$\alpha$	C <sub>G</sub> T G N	Arg Gly Cys Trp
8	$\alpha$	A <sub>C</sub> T T N	Ile Met Leu Val Phe
9	$\alpha$	A <sub>C</sub> G A <sub>C</sub> T G	Lys Thr Met Gln Pro Leu Glu Ala Val
10	$\alpha$	T G G	Trp
11	$\alpha$	N A <sub>G</sub> C	Asn Ser His Arg Asp Gly Tyr Cys
12	$\alpha$	A <sub>T</sub> T C	Ile Phe

Note that any given amino acid appears only once in any set of 4 libraries.

Similar randomisation can now be applied to all three positions:  $\alpha$ ,  $\beta$  and  $\gamma$  of zinc finger proteins, to generate libraries 1-36. In libraries 1-12, the randomisation of residue  $\alpha$  is controlled (in these libraries, residues  $\beta$  and  $\gamma$  are fully randomised - they are specified by the codon NNN). Similarly, libraries 13-24 control the randomisation of position  $\beta$ , and libraries 25-36 control the randomisation of residue  $\gamma$ ).

10

Library	$\alpha$	$\beta$	$\gamma$	Library	$\alpha$	$\beta$	$\gamma$	Library	$\alpha$	$\beta$	$\gamma$
1	A <sup>A</sup> <sub>C</sub> T	N	NNN	13	NNN	A <sup>A</sup> <sub>C</sub> T	NNN	25	NNN	NNN	A <sup>A</sup> <sub>C</sub> T
2	C <sup>A</sup> <sub>C</sub> G	NNN	NNN	14	NNN	C <sup>A</sup> <sub>C</sub> G	NNN	26	NNN	NNN	C <sup>A</sup> <sub>C</sub> G
3	G <sup>A</sup> <sub>C</sub> T	NNN	NNN	15	NNN	G <sup>A</sup> <sub>C</sub> T	NNN	27	NNN	NNN	G <sup>A</sup> <sub>C</sub> T
4	T <sup>A</sup> <sub>C</sub> T	NNN	NNN	16	NNN	T <sup>A</sup> <sub>C</sub> T	NNN	28	NNN	NNN	T <sup>A</sup> <sub>C</sub> T
5	N <sup>A</sup> <sub>C</sub> T	NNN	NNN	17	NNN	N <sup>A</sup> <sub>C</sub> T	NNN	29	NNN	NNN	N <sup>A</sup> <sub>C</sub> T
6	N <sup>C</sup> <sub>C</sub> T	NNN	NNN	18	NNN	N <sup>C</sup> <sub>C</sub> T	NNN	30	NNN	NNN	N <sup>C</sup> <sub>C</sub> T
7	C <sup>G</sup> <sub>C</sub> T	NNN	NNN	19	NNN	C <sup>G</sup> <sub>C</sub> T	NNN	31	NNN	NNN	C <sup>G</sup> <sub>C</sub> T
8	A <sup>G</sup> <sub>C</sub> T	NNN	NNN	20	NNN	A <sup>G</sup> <sub>C</sub> T	NNN	32	NNN	NNN	A <sup>G</sup> <sub>C</sub> T
9	A <sup>A</sup> <sub>C</sub> T	G	NNN	21	NNN	A <sup>A</sup> <sub>C</sub> T	G	33	NNN	NNN	A <sup>A</sup> <sub>C</sub> T
10	T <sup>A</sup> <sub>C</sub> T	G	NNN	22	NNN	T <sup>A</sup> <sub>C</sub> T	G	34	NNN	NNN	T <sup>A</sup> <sub>C</sub> T
11	N <sup>A</sup> <sub>C</sub> T	C	NNN	23	NNN	N <sup>A</sup> <sub>C</sub> T	C	35	NNN	NNN	N <sup>A</sup> <sub>C</sub> T
12	A <sup>T</sup> <sub>C</sub> T	C	NNN	24	NNN	A <sup>T</sup> <sub>C</sub> T	C	36	NNN	NNN	A <sup>T</sup> <sub>C</sub> T

Nucleotide sequences of randomised codons  $\alpha$ ,  $\beta$  and  $\gamma$  in libraries 1-36

Randomisation Strategy A

All 36 gene libraries are expressed to generate zinc finger libraries. These zinc finger libraries are then incubated with a polynucleotide of interest, in such a way as to identify one library from each group of four that binds most strongly to the polynucleotide. For example, each library may be placed in an individual well of a microtitre plate and there incubated with the same trinucleotide.

Consider the controlled randomisation of residue  $\alpha$ . Because in any one group of 4 libraries each amino acid is encoded only once, each amino acid, as residue  $\alpha$ , will occur in only three of the twelve libraries:

09787228.071904  
TOPIC: B2223650

Library	Lys	Asn	Thr	Ile	Met	Gln	His	Pro	Glu	Asp	Ala	Gly	Val	Tyr	Cys	Trp	Phe	Ser	Arg	Leu
1	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	✓	-	✓	-	✓	✓	-	✓	✓	-	✓	-	✓	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	✓	-	-	-	-	✓	-	-	✓	-	✓	-	✓	✓	✓	✓	✓	✓	✓
12	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

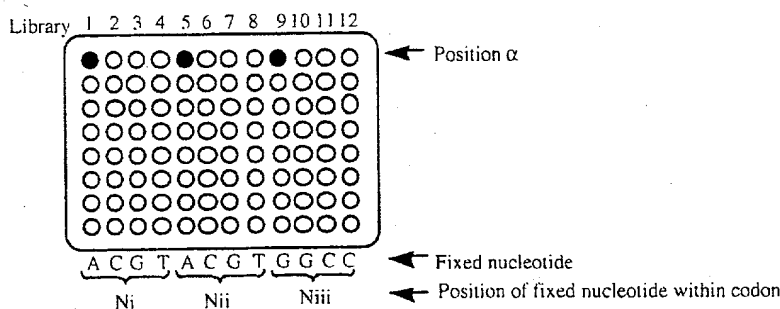
Key:

- ✓: Specified amino acid is present in this library, at position  $\alpha$ .
- : Specified amino acid is **not** present in this library, at position  $\alpha$ .

- 15 -

Presence / absence of an amino acid at position  $\alpha$  within any given library is a direct result of the controlled randomisation and the genetic code.

- This may now be applied to the assay. Consider that libraries 1-12 only are screened with the trinucleotide ATG and that in order for a zinc finger to bind ATG, residue  $\alpha$  must be Lys (lysine). An assay of libraries 1-12 is performed:



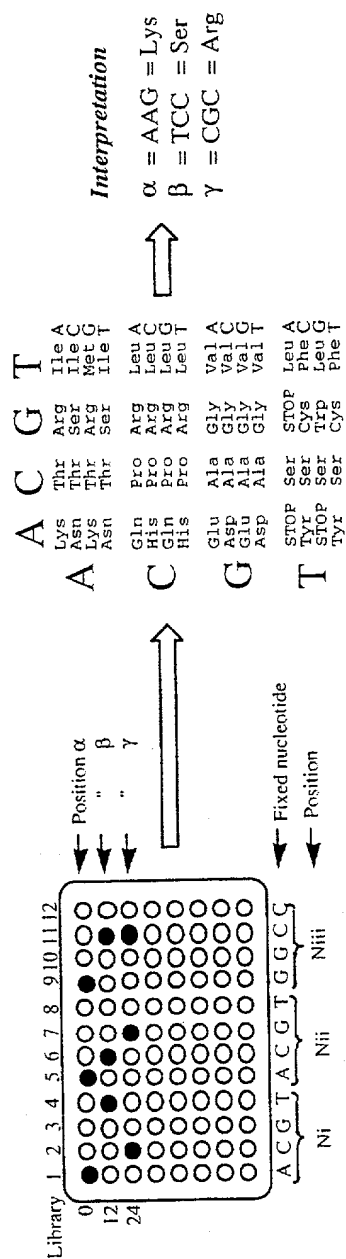
10

- Only libraries 1, 5 and 9 contain lysine as residue  $\alpha$ , therefore only these libraries can emit light. None of the other libraries can emit light, because none of them specify lysine as residue  $\alpha$ . However, this is not the limit of our knowledge. We know the identity of the fixed nucleotide within each library. Moreover, we can read this off directly from the microtitre plate. In this case, the order of fixed nucleotides is AAG.

- Thus, simply from the unique combination of libraries which emit light, we know the genetic code for the amino acid required as residue  $\alpha$ . In this case, the essential fixed nucleotides are AAG, which specifies lysine. We have now linked the genetic code directly to the physical properties of a protein.

20

This principle may be applied to all 36 libraries. In so doing, the genetic codes and thus required identities of all three residues  $\alpha$ ,  $\beta$  and  $\gamma$  will be determined:



- 17 -

This is possible, because in libraries 1-12, residues  $\beta$  and  $\gamma$  are fully randomised. Therefore, in each of libraries 1-12 Ser and Arg are present as residues  $\beta$  and  $\gamma$  within the mixture.

Similarly, when controlled randomisation is applied to residue  $\beta$  (libraries 13-24) residues  $\alpha$  and  $\gamma$  are fully randomised and when controlled randomisation is applied to residue  $\gamma$ , residues  $\alpha$ ,  $\beta$  are fully randomised.

By screening the 36 libraries with each of the 64 trinucleotides, an optimum zinc finger will be found for each trinucleotide. Thus the result is therefore the solution of the zinc finger code whereby DNA binding proteins may now be designed at will.

Should more than three libraries within a given set of twelve produce a signal, then the plates may be washed to remove signals resulting from weak interactions. An end point to the assay has been reached when just three libraries per set of twelve generate a signal.

The above strategy generates libraries of genes which when expressed, yield protein libraries in which two positions are fully randomised and one position has controlled randomisation. In practice, this leads to libraries with between 400 (e.g. library 10) and 3600 (eg. library 9) constituent proteins. These numbers are calculated as follows:

$$\begin{aligned}
 \text{Number of library constituents} &= \text{multiplication of number of possibilities at each position of randomisation} \\
 \text{eg. library 1:} &= \text{position } \alpha \times \text{position } \beta \times \text{position } \gamma \\
 &\quad 5 \quad \times \quad 20 \quad \times \quad 20 \\
 &= \underline{\underline{2000 \text{ constituents (proteins)}}}
 \end{aligned}$$

However, these small libraries result from the degeneracy of the genetic code. In practice, the gene libraries which encode the proteins, randomised as above, will be far larger. For example, again consider

- 18 -

library 1:

Codon	$\alpha$		$\beta$		$\gamma$
Sequence	A A <sub>C</sub> T N		N N N		N N N
Numbers	1x3x4	x	4x4x4	x	4x4x4 = <u>49152 constituents (genes)</u>

The generation of such libraries should not be problematic technically, since libraries far larger than these exist already (eg. Choo and Klug, 1994, PNAS 91, 11163-7). However, it may it may prove beneficial to reduce the gene library sizes to those of the protein libraries. Potential benefits include:

- greater likelihood of full representation within each library (all constituent proteins encoded);
- even representation of each constituent (an equal amount of each constituent protein within a given library);
- consistent optimum codon usage (to maximise expression).

These attributes are desirable because of the degeneracy of the genetic code. Again consider library 1. Within this library, position  $\beta$  is encoded by NNN. When expressed therefore, residue  $\beta$  is 6 times more likely to be serine than it is to be methionine, because serine is encoded six times within NNN for each encoding of methionine.

Such bias within libraries may have an adverse effect on the results of the assay. Any detrimental effect is predicted to be minor - it should occur only if two proteins have similar binding affinities with a given DNA sequence. However, such an eventuality is possible: consider that two zinc fingers with positions  $\alpha$ =Arg,  $\beta$ =Ser,  $\gamma$ =Lys and  $\alpha$ =Arg,  $\beta$ =Met,  $\gamma$ =Lys bind similarly to a given sequence of DNA, with  $\alpha$ =Arg,  $\beta$ =Met,  $\gamma$ =Lys being the optimally binding zinc finger protein. During the assay, the effective concentration of the protein containing serine at position  $\beta$  would be greater than that of the protein containing methionine. Thus, the serine-containing protein might give a stronger signal even though it is not the

FOI 2012-071901



- 19 -

optimum zinc finger for that DNA sequence.

- It may therefore be preferred to substitute the codon MAX for positions of full randomisation (previously NNN), where MAX is a mixture
- 5 containing only the following codons:

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA, GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

- 10 These codons represents those most favoured by *E. coli* for each amino acid (Nakamura et al., (1997), Nucleic Acids Research, 25, 244-245).

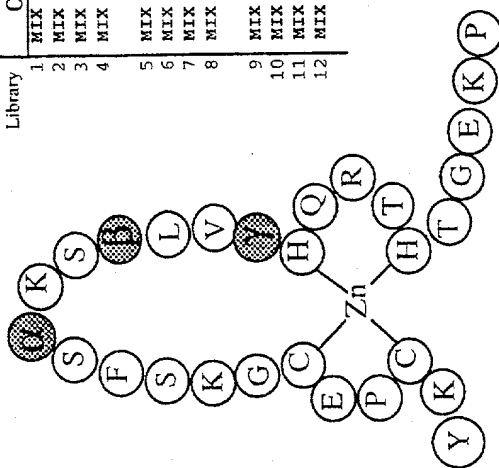
- In order to employ these codons in controlled randomisation, a new division of the codons into sets of 12 libraries is required, as outlined
- 15 in randomisation strategy B:

09/07/2000 10:45:04

Library	$\alpha$	$\beta$	$\gamma$	Library	$\alpha$	$\beta$	$\gamma$
1	MIX 1	M A X	M A X	25	M A X	M A X	MIX 1
2	MIX 2	M A X	M A X	26	M A X	M A X	MIX 2
3	MIX 3	M A X	M A X	27	M A X	M A X	MIX 3
4	MIX 4	M A X	M A X	28	M A X	M A X	MIX 4
5	MIX 5	M A X	M A X	29	M A X	M A X	MIX 5
6	MIX 6	M A X	M A X	30	M A X	M A X	MIX 6
7	MIX 7	M A X	M A X	31	M A X	M A X	MIX 7
8	MIX 8	M A X	M A X	32	M A X	M A X	MIX 8
9	MIX 9	M A X	M A X	33	M A X	M A X	MIX 9
10	MIX 10	M A X	M A X	34	M A X	M A X	MIX 10
11	MIX 11	M A X	M A X	35	M A X	M A X	MIX 11
12	MIX 12	M A X	M A X	36	M A X	M A X	MIX 12

Nucleotide sequences of randomised codons  $\alpha$ ,  $\beta$  and  $\gamma$  in libraries 1-36

*Randomisation strategy B*



wherein mixes 1 to 12 are as detailed in Table 2:

Mix	1	2	3	4	5	6	7	8	9	10	11	12
Codons	AAA AAC ACC AGC ATG ATT	CAG CAT CCG CGC CTG	GAA GAT GCG GGC GTG	TAT TGC TGG TTT	AAA AAC CAG CAT GAA TAT	ACC CCG GCG GAT TGG	AGC CGC GGC GTG TTT	ATG CTG TGG	AAA GAA  CGC GGC TGC	AAC ACC AGC CGC GCG TGC	ATG CAG CCG CTG GCG GCG GTG	ATT CAT GAT TTT TGG
Fixed nucleotide	A	C	G	T	A	C	G	T	A	C	G	T
Position	Ni				Nii				Niii			

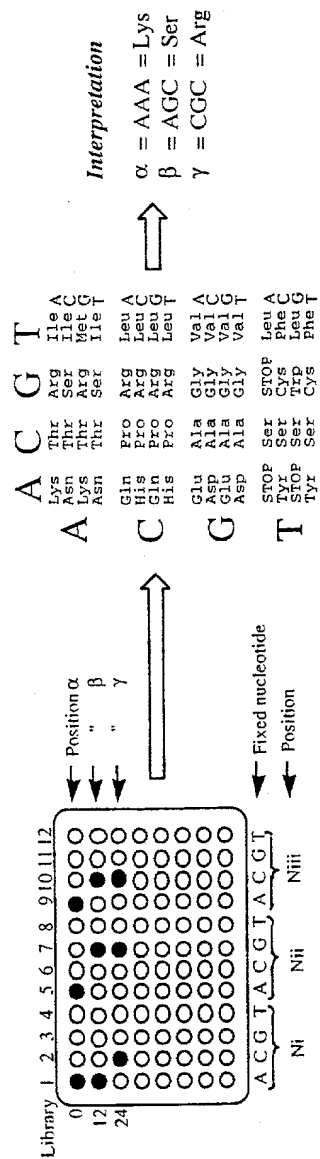
Consider the controlled randomisation of position  $\alpha$  (libraries 1-12). When expressed, position  $\alpha$  will be represented as follows in each library, while positions  $\beta$  and  $\gamma$  are fully randomised:

Library	Lys	Asn	Thr	Ile	Met	Gln	His	Pro	Glu	Asp	Ala	Gly	Val	Tyr	Cys	Trp	Phe	Ser	Arg	Leu
1	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-
2	-	-	-	-	-	✓	✓	✓	-	-	-	-	✓	-	-	-	-	-	✓	-
3	-	-	-	-	-	-	-	-	✓	✓	✓	✓	-	-	-	✓	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	-	-	-
5	✓	✓	-	-	-	✓	✓	-	✓	✓	-	-	-	✓	-	-	-	-	-	-
6	-	-	✓	-	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	✓	✓	-	✓	✓	-
8	-	-	-	✓	✓	-	-	-	-	-	-	-	✓	-	-	-	✓	-	-	✓
9	✓	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-
10	-	✓	✓	-	-	-	-	-	-	-	-	✓	-	-	✓	-	-	✓	-	-
11	-	-	-	-	✓	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-	-
12	-	-	-	✓	-	-	✓	-	-	✓	-	-	-	✓	-	-	-	-	-	-

- 23 -

The changes in controlled randomisation will affect the library numbers which produce a signal and therefore the interpretation of the assay results. However, the principles of controlled randomisation and the mechanism of assay interpretation remain unchanged. Using randomisation strategy B, the example illustrated above is reiterated:

09787238.071901



Note the different fixed nucleotides in libraries 9-12 and that different libraries now light up. The end result: α=Lys, β=Ser, γ=Arg is the same, however.

Randomisation strategy A is in principle, the easier strategy to implement technically. However, strategy B is preferred. Gene libraries of much smaller size are required. Although construction of these highly-controlled libraries is technically demanding, it is much more likely that the libraries encode all required proteins and moreover that those proteins are encoded in similar proportions, so removing potential difficulties in the SPA library assays.

Construction of these gene libraries may be achieved by cloning oligonucleotide cassettes between two appropriately positioned restriction sites which flank positions  $\alpha$  and  $\gamma$ . Construction of the oligonucleotide cassettes requires a set of sixty-one oligonucleotides comprising one fully-randomised "template" oligonucleotide and three pools of selection oligonucleotides. The template oligonucleotide is of sequence

3'-----NNN-----NNN-----NNN-----5'

where "-----" represents the invariant DNA and NNN the positions of randomisation within the non-coding strand of the gene. The intervening sequences "-----" are conveniently between 3 and 21 bases in length.

The pools of selection oligonucleotides contain twenty individual oligonucleotides of sequence

Lys: 5'-----AAA-----3'  
 Asn: 5'-----AAC-----3'  
 Thr: 5'-----ACC-----3'  
 Ser: 5'-----AGC-----3'  
 Met: 5'-----ATG-----3'  
 Ile: 5'-----ATT-----3'  
 Gln: 5'-----CAG-----3'  
 His: 5'-----CAT-----3'  
 Pro: 5'-----CCG-----3'

- 26 -

Arg: 5'-----CGC-----3'  
 Leu: 5'-----CTG-----3'  
 Glu: 5'-----GAA-----3'  
 Asp: 5'-----GAT-----3'  
 5 Ala: 5'-----GCG-----3'  
 Gly: 5'-----GGC-----3'  
 Val: 5'-----GTG-----3'  
 Tyr: 5'-----TAT-----3'  
 Trp: 5'-----TGG-----3'  
 10 Cys: 5'-----TGC-----3'  
 Phe: 5'-----TTT-----3'

where the sequence "-----" is of suitable length and base sequence to  
 base pair with the non-variant regions of the template and the defined  
 15 codon corresponds to one of those comprising the "MAX" set of codons  
 (defined herein at page 18, line 5). The defined codon corresponds to a  
 position of randomisation and must be either at or near to one end of the  
 oligonucleotide. A complete selection pool represents a set of twenty such  
 oligonucleotides, in order that all codons contained within "MAX" are  
 20 represented and all twenty amino acids are encoded.

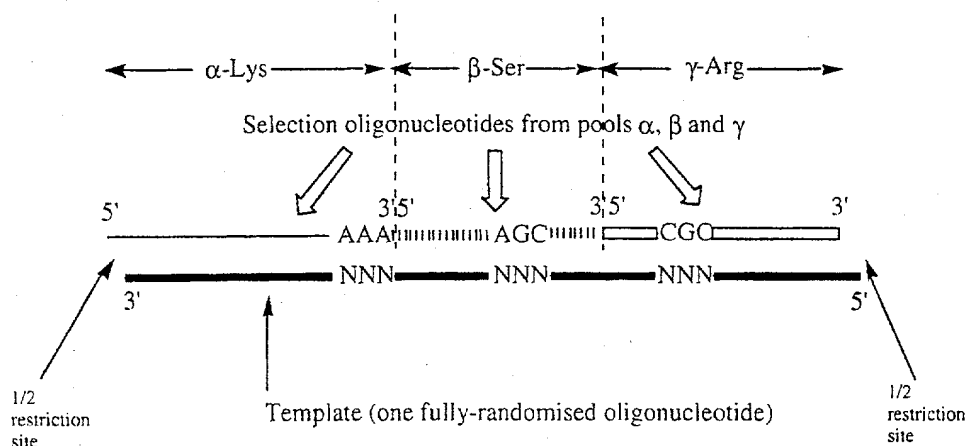
The invention enables fully randomised libraries, positionally  
 fixed libraries and individual genes to be constructed. Oligonucleotides  
 encoding the required amino acid at each position of randomisation would  
 be taken from each selection pool. For example, if full randomisation is  
 25 required at a given position, then all 20 selection oligonucleotides would be  
 taken. If positional fixing were required, then all oligonucleotides where the  
 "MAX" codon begins with A (for example) would be taken. If a single amino  
 acid were required at the position of randomisation, the single selection  
 oligonucleotide corresponding to that amino acid would be taken.

09787228-071901



### Construction of a single zinc finger gene encoding $\alpha$ =Lys, $\beta$ =Ser, $\gamma$ =Arg

The selection oligonucleotides  $\beta$ -Ser and  $\gamma$ -Arg are treated with T4 polynucleotide kinase and ATP in order to attach 5' phosphate groups and so enable them to participate in ligation reactions. These two oligonucleotides, together with the selection oligonucleotide  $\alpha$ -Lys and the template oligonucleotide are combined, heated to 90°C and allowed to cool slowly to room temperature, in order to allow complementary sequences of DNA to base pair as shown below:



#### KEY:

- Invariant DNA sequence within pool  $\alpha$
- ||||| Invariant DNA sequence within pool  $\beta$
- Invariant DNA sequence within pool  $\gamma$
- Invariant DNA sequence of the template oligonucleotide

The resulting oligonucleotide cassette is then inserted into the appropriate restriction sites in the zinc finger gene, so generating the zinc finger gene  $\alpha$ =Lys,  $\beta$ =Ser,  $\gamma$ =Arg. None of the other sequences contained in the template oligonucleotide are cloned, since only the double stranded DNA  
 5 cassette will be ligated into the parental gene. Selection from the template oligonucleotide is thus achieved by addition of the three selection oligonucleotides.

### Construction of zinc finger library 1

10 The selection oligonucleotides  $\beta$ -MAX and  $\gamma$ -MAX (where MAX = an entire selection pool) are treated with T4 polynucleotide kinase and ATP in order to attach 5' phosphate groups and so enable them to participate in ligation reactions. These two oligonucleotide pools, together with the selection oligonucleotide  $\alpha$ -MIX 1 where MIX 1 is the following  
 15 mixture of oligonucleotides:

	$\alpha$ -Lys:	5'-----AAA-----3'
	$\alpha$ -Asn:	5'-----AAC-----3'
	$\alpha$ -Thr:	5'-----ACC-----3'
20	$\alpha$ -Ser:	5'-----AGC-----3'
	$\alpha$ -Met:	5'-----ATG-----3'
	$\alpha$ -Ile:	5'-----ATT-----3'

and the template oligonucleotide are combined, heated to 90° C and  
 25 allowed to cool slowly to room temperature, in order to allow complementary sequences of DNA to base pair as above.

The resulting mixture of oligonucleotide cassettes is then inserted into the appropriate restriction sites in the zinc finger gene, so generating the zinc finger library 1. None of the other sequences contained  
 30 in the template oligonucleotide are cloned, since only the double stranded DNA cassettes will be ligated into the parental gene. Selection from the

template oligonucleotide is thus achieved by addition of the three pools of selection oligonucleotides. Note that the number of genes exactly matches the number of encoded proteins and that no truncated proteins should result, since "MAX" contains no termination codons.

5

### **Generalised application to randomised peptides**

The above technique may also be used to generate genes encoding fully randomised peptides, without intervening conserved gene sequences. Again, the number of genes will exactly match the number of encoded peptides. In the case of a fully randomised peptide library without positional fixing, just 21 oligonucleotides are required: a fully-randomised template oligonucleotide of the desired length and a set of the twenty "MAX" trinucleotides. Annealing between the set of "MAX" trinucleotides and the template will generate cassettes encoding all possible peptides, dependent on complete representation within the template oligonucleotide, which will decrease with oligonucleotide length.

Positionally fixed, random peptides may be made similarly, although a set of twelve templates will be required for each codon. Here, for a given codon, the non-coding template strand will be fixed alternatively as T, G, C and A at each nucleotide and the "MAX" trinucleotides annealed as above.

a) The above strategies A and B involve designing sets of libraries of genes which in turn may be expressed to generate corresponding libraries of proteins.

The method of the invention involves incubating a set of libraries of proteins with a specific binding partner, observing specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner. Although other assay techniques are possible, this method is preferably performed using scintillation proximity assay (SPA) technology. Briefly, this technology involves providing a support which comprises a

scintillant which emits light when subjected to electrons (e.g.  $\beta$  particles) or other forms of radiation resulting from decomposition of a radioisotope.

The support may be massive, e.g. the base of each well of a microtitre plate, or may be particulate. One assay reagent is immobilised on the support. Another assay reagent is radiolabelled and is partitioned between two fractions, one bound to the support and the other free in solution. The relative size of the two fractions is arranged to be related to the presence or the concentration of an analyte of interest. The radioisotope is chosen such that reagent bound to the support causes the scintillant in the support to emit light, while reagent free in solution does not (on account of the short mean free path of the radiation) significantly affect the scintillant substance.

Various assay formats are possible. For example, each library of a set of libraries can be immobilised in an individual well, either of a standard microtitre plate or of a scintillant containing microtitre plate. A specific binding partner of the proteins is labelled and introduced into each well. Labels can be radiometric, luminescent, for example fluorescent or may be enzyme. Where radiometric or luminescent labels are used, a specific binding interaction can be investigated in real time. Where enzyme labels are used the interaction can be investigated upon the addition of the appropriate reagents needed to generate a signal. Where several wells emit a signal, repeated washing can be used to remove weakly interacting species until the specific binding partner remains bound only in a single well. This ability to identify a single library (as opposed to a small pool of libraries) that bind most strongly to any particular specific binding partner, is a valuable feature, and an advance on assay techniques used previously for similar purposes.

Alternatively, the specific binding partner can be immobilised in each well of the SPA microtitre plate. Each protein library is radiolabelled and introduced into a different well of the plate for interaction with the specific binding partner. Alternative assay formats, in which neither the protein library nor its specific binding partner, but rather a third

reagent is radiolabelled, are well known in the art.

Techniques for immobilising protein or other assay reagents on SPA surfaces in forms suitable for taking part in SPA assays, are well known in the art. Development of suitable techniques should not amount to more than the routine optimisation ordinarily required for assays of this kind. Detection of interactions by non-radioactive assay and imaging techniques such as luminescent, for example fluorescent, detection or colorimetric detection of interactions between, for example, biotin linked and streptavidin linked partners is also envisaged.

Most zinc finger proteins form the DNA recognition module of transcription factors, which serve to switch genes on or off. Already, several examples exist where novel transcription factors have been engineered, by changing their zinc fingers (Choo *et al* (1994), Nature 372, 642-5). Similarly, zinc fingers have been linked to restriction endonuclease cleavage domains, to generate novel restriction endonucleases (e.g. Kim *et al* (1996), PNAS 93, 1156-60). The application of zinc fingers is almost limitless - when ever a need arises to link something to a specific sequence of DNA, it can be met with a series of zinc fingers. However, in order to design DNA-binding proteins at will, there must be available one zinc finger for each trinucleotide. This invention provides enabling technology to achieve that object.

### Example

The example involves a single protein, comprising three zinc fingers. Controlled randomisation is applied only to the central zinc finger. The two outer zinc fingers are present simply to ensure correct registry with the target DNA sequence and to increase overall binding strength (Choo and Klug, (1994) PNAS 91, 11163-67; Berg (1997) Nature Biotech. 15, 323).

The work is divided into four stages: gene synthesis, gene expression, radiometric and colorimetric assay formats, assay results and

proof of principle.

### Gene Synthesis:

A gene was designed and synthesised to encode the protein

5 (SEQ ID NO: 1)

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H

10

T G E K P Y K C P E C G K S F S K K S H L V A H Q R T H.

#### KEY:

X linker residues

15 X zinc co-ordinating residues

X DNA-contacting residues ( $\alpha$ ,  $\beta$  and  $\gamma$ ) (positions -1, +3 and +6)

This protein corresponds to three repeats of Berg's

20 consensus zinc finger sequence (Krizek *et al.*, (1991) JACS 113, 4518-23),  
with DNA-contacting residues from the first zinc finger of transcription  
factor Sp1 (Berg (1992) PNAS 89, 11109-10; Shi and Berg, (1995) Chem  
& Biol. 2, 83-89). Each zinc finger sequence is preceded by a *Kruppel*-type  
linker peptide (Choo and Klug (1993) NAR 21, 3341-6). By analogy to  
25 previous precedent (Shi and Berg, 1995), the three repeats of this novel  
zinc finger peptide are expected to bind to the dsDNA sequence  
5'-GGG GGG GGG-3'.

To maximise gene expression, on converting the sequence  
into DNA, *E. coli* codon preference was employed (Wada *et al.* (1992)

30 NAR20 sup., 2111-8). Wherever possible, first preference codons were  
used. However, in some instances, second preference codons were also  
employed. These limited sequence repetition within the gene, necessary to  
prevent potential intragenic recombination events, which would be

deleterious to ensuing experiments. In practice, a maximum repeat length of 8 base pairs was mostly achieved. Use of second preference codons also allowed the incorporation of restriction enzyme sites within the gene. The final gene sequence, restriction sites and codon usage are illustrated in Figure 1.

### Gene Expression

In the current assay format, the zinc finger gene is fused to the glutathione-S-transferase gene in the vector pGEX2TK (Amersham Pharmacia Biotech). Expression of this construct leads to a 36.5 kD protein comprising GST at the amino terminus and the zinc finger protein at the carboxyl terminus. Gene expression is performed in *E. coli* BL21 cells according to manufacturer's instructions. The resulting fusion protein is then purified using glutathione-Sepharose (Amersham Pharmacia Biotech) according to manufacturer's instructions. Use of the pGEX2TK vector allows for the subsequent radiolabelling of the protein if required.

### Assay formats for assessing zinc finger – DNA interactions

- 20 Direct attachment of GST fusion protein to microtitre plates, followed by colorimetric detection of biotinylated DNA (Assay format 1)

GST or GST ZF protein (4 pmoles per well) was immobilised in microtitre wells in carbonate buffer, pH 9.2, for 18 hrs. The plates were washed three times in TBS-Tween (0.3% Tween) and then blocked in the same buffer for 3 hrs. After washing, 2-fold serial dilutions of DNA were added to each well. The protein and DNA were incubated together for 2 hrs at room temperature, and the wells were then washed 3 times in TBS-Tween. As negative controls, experiments were performed in the absence of DNA, to assess binding of GST / GST ZF proteins by the streptavidin conjugate. Bound DNA was detected by adding streptavidin / peroxidase conjugate, which was removed by 3 washes in TBS. Finally, the conjugate

was detected colorimetrically according to manufacturer's instructions. All reactions were performed in duplicate. Figure 1 demonstrates that interaction between the zinc finger protein and its target DNA sequence may be assessed using this assay format. In figures 1, 2 and 3, the legend 'bkg' denotes background detection levels.

Direct attachment of GST fusion protein to microtitre plates, followed by scintillation-based detection of radiolabelled DNA (Assay format 2)

GST or GST ZF protein (4 pmoles per well) was immobilised in microtitre wells in carbonate buffer, pH 9.2, for 18 hrs. The plates were washed three times in TBS-Tween (0.3% Tween) and then blocked in the same buffer for 3 hrs. After washing, 2-fold serial dilutions of radiolabelled DNA were added to each well. The protein and DNA were incubated together for 2 hrs at room temp, and the wells were then washed 3 times in TBS-Tween. Bound DNA was detected by scintillation counting. All reactions were performed in duplicate. Figure 2 demonstrates that interaction between the zinc finger protein and its target DNA sequence may be assessed using this assay format.

Antibody-based attachment of GST fusion protein to microtitre plates, followed by scintillation-based detection of radiolabelled DNA (Assay format 3)

One µg of protein A was attached to the surface of each microtitre well in carbonate buffer, pH 9.2, for 18 hrs. The plates were washed three times in TBS-BSA (2% BSA) and then blocked in the same buffer for 3 hrs. Anti-GST antibody (1µg) was added to each well in the same buffer and incubated at room temperature with rocking, for 1 hr. The plates were washed 3 times in TBS-BSA and then incubated for 1 hr with 4 pmoles GST / GST ZF protein per well. After washing away unbound protein, the plates were incubated for 2 hrs at room temp with 2-fold serial dilutions of radiolabelled DNA. Unbound DNA was removed by 3 washes



in TBS-BSA. As negative controls, experiments were performed in the absence of antibody, to assess any binding of radiolabelled DNA by protein

A. All reactions containing GST / GST ZF were performed in duplicate.

Figure 3 demonstrates that interaction between the zinc finger protein and  
5 its target DNA sequence may be assessed using this assay format.

### Conclusion

Three adsorption-based assay formats have been developed.

All assay formats demonstrate interaction between the protein and its DNA  
10 target sequence. In each case, the protein is immobilised and the DNA is  
in solution. Labelled DNA is bound by the immobilised protein and then  
detected according to the nature of the label. Radiolabelled DNA is  
detected using scintillation-based methods or appropriate imaging  
technology. Non-radiometrically labelled DNA is detected using  
15 colorimetric techniques and a spectrophotometer. The assay formats are  
also applicable to fluorescently labelled DNA, where imaging technology  
would be used to detect the bound DNA.

09787229-071901

- 36 -

## CONFIRMATION

CLAIMS

- 5 1. A set of libraries of genes which code for proteins which are capable of specific binding interactions with a specific binding partner by amino acid residues at at least two specified positions including a first specified position and at least one other specified position, which set of libraries consists of:
- 10 a) 6 to 20 libraries in which each library has a triplet that codes for at least one but less than 20 amino acids at said first specified position, and is randomised at the or each triplet coding for the said at least one other specified position, the arrangement being such that interactions of the proteins coded for by the said 6 to 20 libraries with a specific binding
- 15 partner identifies a triplet that codes for an amino acid at the said first specified position that takes part in the specific binding interaction, and
- b) 6 to 20 libraries in each of which libraries said first specified position is randomised and a different one of said at least one other specified positions has a triplet that codes for at least one but less than 20
- 20 amino acids.
2. The set of libraries of genes as claimed in claim 1, which set of libraries consists of:
- a) 12 libraries in which each library has a triplet that codes for
- 25 one or several but less than 20 amino acids at the said first determined position, the triplets being as shown in Table 1 or Table 2, and
- b) 12 libraries of corresponding design for each of the said one or more other determined positions.
- 30 3. The set of libraries of genes as claimed in claim 1 or claim 2, wherein the genes code for zinc fingers.

## CONFIRMATION

- 37 -

4. The set of libraries of genes as claimed in claim 3, which set consists of 36 libraries in three groups of 12 libraries which code for amino acids at the -1 and +3 and +6 positions respectively.

5. The set of libraries of genes as claimed in claim 3 or claim 4, wherein each gene codes for a protein comprising 3 zinc fingers.

6. The set of libraries of genes as claimed in claim 5, wherein each gene codes for a protein having the sequence (SEQ ID NO: 2)

10

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

15

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H.

where X is any amino acid

7. A set of libraries of proteins, which proteins are capable of specific binding interactions with a specified binding partner by amino acid residues at at least one specified position including a first specified position and at least one other specified position, which set of libraries consists of:

- a) 6 to 20 libraries in which each library has at least one but less than 20 amino acid residues at the said first specified position and is randomised at the said at least one other determined position, the arrangement being such that interaction of the 6 to 20 libraries with a specific binding partner identifies an amino acid residue at the said first specified position that takes part in the specific binding interaction, and
- b) 6 to 20 libraries in each of which libraries said first specified position is randomised and a different amino acid is present at at least one other specified position.

## CONFIRMATION

- 38 -

8. The set of libraries of proteins as claimed in claim 7, which set of libraries consists of

a) 20 libraries in which each library has one specified amino acid residue at the said first determined position and is randomised at the said

5 one or more other determined positions, and

b) 20 libraries of corresponding design for each of the said one or more other determined positions.

9. The set of libraries of proteins as claimed in claim 7 or claim  
10 8, wherein the proteins are zinc fingers.

10. The set of libraries of proteins as claimed in claim 7, which set consists of 60 libraries in three groups of 20 libraries with specified amino acids at the -1 and +3 and +6 positions respectively.

15 11. The set of libraries of proteins as claimed in claim 9 or claim 10, wherein each protein comprises three zinc fingers.

12. The set of libraries of proteins as claimed in claim 11, wherein  
20 each protein has the sequence (SEQ ID NO: 2)

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H

25 T G E K P Y K C P E C G K S F S X K S X L V X H Q R T H.

where X is any amino acid

30 13. A set of libraries of genes which code for the set of libraries of proteins defined in any one of claims 7 to 12.

CONFIRMATION

- 39 -

14. A method of identifying a protein which interacts with a specific binding partner, which method comprises providing a set of libraries of proteins as defined in any one of claims 7 to 12, incubating the specific binding partner with each library of the set, observing specific binding interactions with certain libraries of the set, and using the observations to identify a protein which interacts with the specific binding partner.
15. The method as claimed in claim 14, wherein the specific binding partner is a polynucleotide.
16. The method as claimed in claim 14, wherein the specific binding interactions are observed by radiometric or luminescent assay.
17. The method as claimed in claim 14, wherein the specific binding interactions are observed by imaging means.
18. The method as claimed in claim 14, wherein the specific binding interactions are observed by scintillation proximity assay.
19. The method as claimed in claim 18, wherein the sets of libraries of proteins are immobilised on scintillation proximity assay surfaces and the specific binding partner is radiolabelled.
20. The method of claim 18 or claim 19, wherein after incubation the scintillation proximity assay surfaces are washed to distinguish stronger specific binding interactions from weaker ones.

## CONFIRMATION

- 40 -

21. The method as claimed in claim 14, wherein the specific binding interactions are observed by colorimetric means.

22. The method as claimed in claim 21, wherein the specific binding partner is biotinylated and the specific binding interaction is detected using a signal generating streptavidin conjugate.

23. The method as claimed in claim 21 or claim 22 wherein after incubation the binding interactions are washed to distinguish stronger specific binding interactions from weaker ones.

24. A protein having the sequence (SEQ ID NO: 1)

T G E K P Y K C P E C G K S F S . K S + L V \$ H Q R T H

T G E K P Y K C P E C G K S F S . K S + L V \$ H Q R T H

T G E K P Y K C P E C G K S F S . K S + L V \$ H Q R T H.

25. A gene which codes for the protein of claim 24.

26. A method of constructing randomised gene libraries in which the number of genes is the same as the number of encoded proteins and which contain no termination codons at the predetermined positions of randomisation, the method comprising the steps of:

- a) providing a template oligonucleotide which is fully randomised at predetermined codon positions;
- b) for each predetermined codon position providing a pool of selection oligonucleotides, wherein each member of said pool contains a different codon selected from the group consisting of

CONFIRMATION

- 41 -

AAA, AAC, ACC, AGC, ATG, ATT, CAG, CAT, CCG, CGC, CTG, GAA,  
GAT, GCG, GGC, GTG, TAT, TGG, TGC, TTT.

at the predetermined codon position;

- 5 c) selecting one or more selection oligonucleotides from each  
pool in order to encode the required gene or library;
- d) allowing the ligated selected oligonucleotides from each pool  
to hybridise with the template oligonucleotide;
- e) forming one or more constructs by ligating the hybridised  
10 selection oligonucleotides together;
- f) removing a region from a gene of interest corresponding to  
the hybridised product from step e);
- g) forming a gene library or genes by ligating the products  
from step e) into the said gene of interest wherein the said gene of interest  
15 is contained within a suitable expression vector.

27. A method of producing proteins encoded by the randomised  
gene libraries of claim 26 comprising the steps of:

- a) transforming a suitable host cell with the gene or gene  
20 library of claim 27 construct;
- b) expressing the genes to form proteins;
- c) purifying the proteins.

1/2

Fig.1.

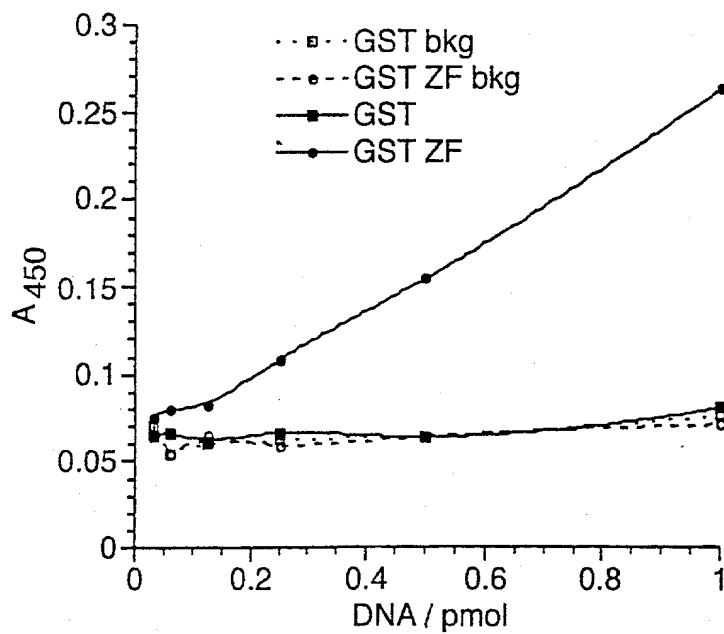


Fig.2.

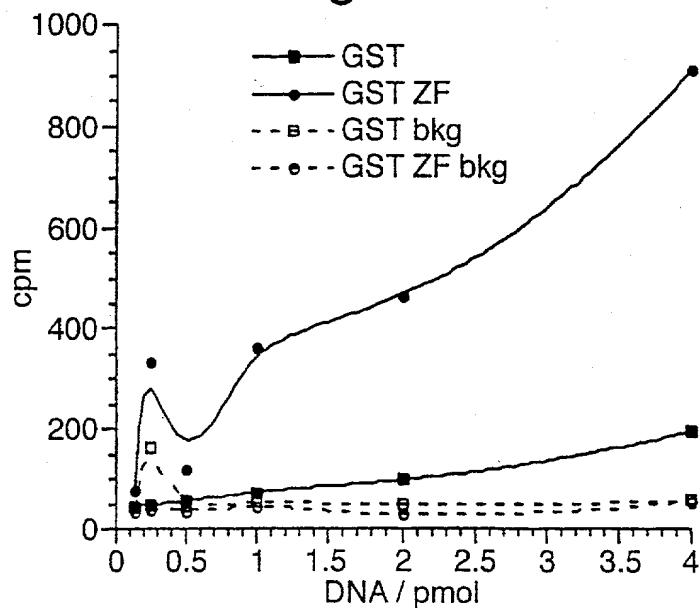
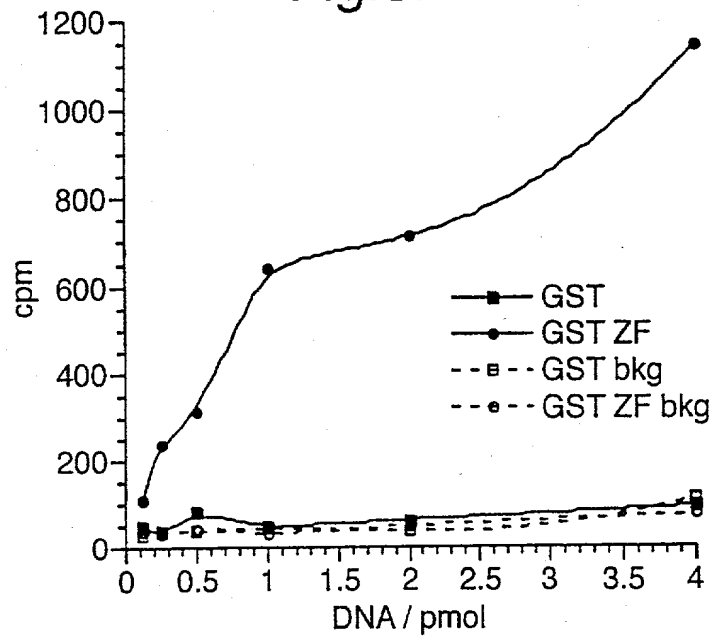




Fig.3.



Please type a plus sign (+) inside this box → ☐

PTO/SB/01 (12-97)  
Approved for use through 9/30/00. OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

<b>DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63)</b>	<b>Attorney Docket Number</b>	PL-9830
	<b>First Named Inventor</b>	A. Hine
	<b>COMPLETE IF KNOWN</b>	
	<b>Application Number</b>	09 / 787,228
	<b>Filing Date</b>	14-Mar-2001
	<b>Group Art Unit</b>	To be assigned
<input type="checkbox"/> Declaration Submitted with Initial Filing	<b>OR</b>	<input checked="" type="checkbox"/> Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)
<b>Examiner Name</b>		To be assigned

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

Gene and Protein Libraries and Methods Relating Thereto

the specification of which

(Title of the Invention)

☐ is attached hereto  
OR

☒ was filed on (MM/DD/YYYY) 03/14/2001 as United States Application Number or PCT International

Application Number 09/787,228 and was amended on (MM/DD/YYYY) (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
98307434.5	EP	09/14/1998	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (MM/DD/YYYY)	<input type="checkbox"/> Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

[Page 1 of 2]

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Please type a plus sign (+) inside this box → ☐

PTO/SB/01 (12-97)  
Approved for use through 9/30/00. OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

## DECLARATION — Utility or Design Patent Application

I hereby claim the benefit under 35 U.S.C. 120 of any United States application(s), or 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT international application in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (if applicable)
PCT/GB99/03081	09/14/1999	

☐ Additional U.S. or PCT international application numbers are listed on a supplemental priority data sheet PTO/SB/02A attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business with the Patent and Trademark Office connected therewith:

☒ Customer Number 22840

OR

☐ Registered practitioner(s) name/registration number listed below

Place Customer Number Here  
22840

Name	Registration Number	Name	Registration Number

☐ Additional registered practitioner(s) named on supplemental Registered Practitioner Information sheet PTO/SB/02C attached hereto.

Direct all correspondence to: ☒ Customer Number 22840 OR ☐ Correspondence address below

Name			
Address			
Address			
City	State	ZIP	
Country	Telephone	Fax	

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle (if any))		Family Name or Surname	
Anna Victoria		Hine	
Inventor's Signature	<i>Anna V. Hine</i>	Date	2 Apr 01
Residence: City	State	Country	Citizenship
		GB	GB
Post Office Address			
Bracebridge House, 139 Chester Road South			
Post Office Address			
Kiddeminster, Worcestershire DY10 1XB Great Britain			
City	State	ZIP	Country

☒ Additional inventors are being named on the 1 supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto

Please type a plus sign (+) inside this box → ☐

PTO/SB/02A (3-97)  
Approved for use through 9/30/98. OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

## DECLARATION

ADDITIONAL INVENTOR(S)  
Supplemental Sheet  
Page 1 of 1

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Jane Leonie

Morgan

Inventor's  
Signature

Date

11 APR 01

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

197 Litchfield Street, Frazzley Tamworth

WREN COTTAGE,  
1 NEWBURY ROAD,

Post Office Address

Staffordshire B78 3QF  
WHITECHURCH, HAMPSHIRE, RG28 7DY,

Great Britain

GBY

City

State

ZIP

Country

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Albert Francis

Santos

Inventor's  
Signature

Date

27 MAR 01

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

Llwynceilyn, Cefn Bychan, Pentyrch

Post Office Address

Cardiff, Mid Glamorgan CF15 9PG Great Britain

GBY

City

State

ZIP

Country

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

David

Palfrey

Inventor's  
Signature

Date

3rd APR 01

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

110 George Street, New Arley, Nuneaton

Post Office Address

Warwickshire CV7 8GP Great Britain

GBY

City

State

ZIP

Country

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Please type a plus sign (+) inside this box → ☐

PTO/SB/02A (3-97)  
Approved for use through 9/30/98. OMB 0651-0032  
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

## DECLARATION

ADDITIONAL INVENTOR(S)  
Supplemental Sheet  
Page 1 of 1

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Jane Leonie

Morgan

Inventor's  
Signature

*J. Morgan*

Date

07/06/01

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

Wren Cottage, 1 Newbury Road

Post Office Address

Whitchurch, Hampshire, RG28 7DY Great Britain

City

State

ZIP

Country

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

Albert Francis

Santos

Inventor's  
Signature

Date

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

Llwynceilyn, Cefn Bychan, Pentyrch

Post Office Address

Cardiff, Mid Glamorgan CF15 9PG Great Britain

City

State

ZIP

Country

Name of Additional Joint Inventor, if any:

☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])

Family Name or Surname

David

Palfrey

Inventor's  
Signature

Date

Residence: City

State

Country

GB

Citizenship

GB

Post Office Address

110 George Street, New Arley, Nuneaton

Post Office Address

Warwickshire CV7 8GP Great Britain

City

State

ZIP

Country

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

09/787228

JC02 Rec'd PCT/PTO 1 4 MAR 2001

## SEQUENCE LISTING

<110> Hine, Anna V  
 Morgan, Leonie J  
 Santos, Albert F  
 Palfrey, David

<120> Gene and Protein Libraries and Methods Relating Thereto

<130> PL9830

<140> To be assigned

<141> 2001-03-14

<150> PCT/GB99/03081

<151> 1999-09-14

<150> 98307434.5 EP

<151> 1998-09-14

<160> 2

<170> PatentIn Ver. 2.1

<210> 1

<211> 28

<212> PRT

<213> synthetic construct

<400> 1

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser  
 1 5 10 15

Lys Lys Ser His Leu Val Ala His Gln Arg Thr His  
 20 25

<210> 2

<211> 28

<212> PRT

<213> synthetic construct

<400> 2

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser  
 1 5 10 15

Xaa Lys Ser Xaa Leu Val Xaa His Gln Arg Thr His  
 20 25

09/787228 071901

## SEQUENCE LISTING

<110> Hine, Anna V  
 Morgan, Leonie J  
 Santos, Albert F  
 Palfrey, David

<120> Gene and Protein Libraries and Methods Relating Thereto

<130> PL9830

<140> To be assigned  
 <141> 2001-03-14

<150> PCT/GB99/03081  
 <151> 1999-09-14

<150> 98307434.5 EP  
 <151> 1998-09-14

<160> 2

<170> PatentIn Ver. 2.1

<210> 1  
 <211> 28  
 <212> PRT  
 <213> synthetic construct

<400> 1  
 Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser  
 1 5 10 15  
 Lys Lys Ser His Leu Val Ala His Gln Arg Thr His  
 20 25

<210> 2  
 <211> 28  
 <212> PRT  
 <213> synthetic construct

<400> 2  
 Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser  
 1 5 10 15  
 Xaa Lys Ser Xaa Leu Val Xaa His Gln Arg Thr His  
 20 25

09787228-031901